

# Ragini Gupta

HPC\_DES

## Title and Abstract Submission for HPC Showcase

LA-UR- 20-25729

Title: [Deploying Machine Learning Workflows into HPC environment](#)

**Abstract:** Modern high performance computing (HPC) applications require complex workflows for running scientific simulation and big data analytics tasks across cluster nodes. With the increasing demand to solve more complicated problems and processing large amount of data in HPC, machine learning is gaining a large momentum. Most machine learning algorithms are focused on utilizing high performance technologies that help boost the performance of big data and data mining frameworks. Typically, on a front-end node of HPC cluster, a workflow intertwined with dependent jobs is submitted which follows a strict, historic process. These jobs are distributed across thousands of compute nodes and once the job finishes the output from these nodes is aggregated into an output file rendered back to the user. In this project, we explored a new adaptable workflow for running machine learning jobs that are intertwined in a loosely coupled fashion to execute different machine learning stages in any user-specific application. An open standard workflow specification language, Common Workflow Language (CWL) is employed to define the pipeline framework for an end-to-end machine learning algorithm by expressing workflow in a portable way combining disparate command line tools for different jobs and passing files around in top-to-bottom scheme. These workflows can be further built with HPC's portable container environment called, Build and Execution Environment (BEE). Since machine learning algorithms are driven by tuning parameters, these parameters are embedded as arguments in command line tool descriptions to achieve accurate and optimized results. The project provides a proof of concept for executing machine learning algorithms as a workflow with BEE for a given HPC application.